(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 May 2002 (30.05.2002)

PCT

(10) International Publication Number
WO 02/43360 A2

(51) International Patent Classification⁷: H04M 3/00

(21) International Application Number: PCT/US01/45171

(22) International Filing Date: 1 November 2001 (01.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/244,651 1 November 2000 (01.11.2000) US

(71) Applicant: LPS ASSOCIATES, LLC [US/US]; 809 Bayridge Drive, Gaithersburg, MD 20878 (US).

(72) Inventors: STINGLEY, Patrick; 8807 Second Avenue, Silver Spring, MD 20910 (US). MOHNKERN, Scott, A.; 809 Bayridge Drive, Gaithersburg, MD 20878 (US). SOLOMON, Louis, P.; 213 Hart Road, Gaithersburg, MD 20878 (US). MORENOFF, Jerry; 9614 Weathered Oak Drive, Bethesda, MD 20817 (US).

(74) Agent: KNOPS, Peter, C.; Lathrop & Gage LC, Suite 2800, 2345 Grand Boulevard, Kansas City, MO 64108 (US).

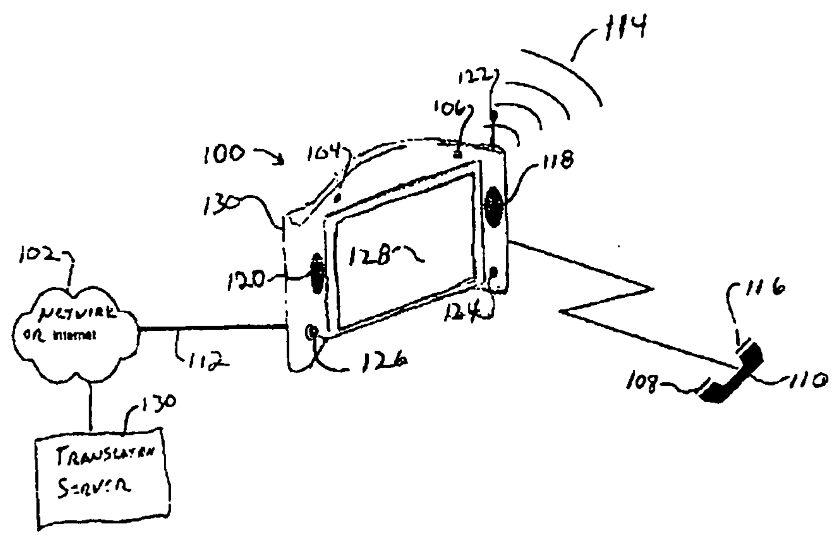
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) Title: MULTIMEDIA INTERNET MEETING INTERFACE PHONE



(57) Abstract: A teleconferencing system is manufactured using components that assure use of the system as a dedicated Internet protocol appliance having multimedia capability. Audio and video data are processed using packetization for separate data streams that can be adjusted to give the audio data a priority which avoids unpleasant audio latency effects. The size of data packets can be adjusted to optimize network transmission capacities by achieving a balance between error transmission rates requiring retransmission of data and data packet sizes. Data compression algorithms include wavelet compression methods.

WO 02/43360 A2

WO 02/43360 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

MULTIMEDIA INTERNET MEETING INTERFACE PHONE

RELATED APPLICATIONS

This applications claims benefit of priority to provisional application serial number 60/244,651 filed November 1, 2000, which is hereby incorporated by reference to the same extent as though fully disclosed herein.

BACKGROUND OF THE INVENTION

1. Field of The Invention

The present invention pertains to method and apparatus for teleconferencing. More particularly, a dedicated network appliance is adapted for specialized teleconferencing purposes through the use of use of an embedded processor and compression algorithms to provide robust audio and/or video teleconferencing capabilities.

2. **Description of the Related Art**

In a historical context, teleconferencing devices involving telephone and computer technologies have evolved independently of one other. Computer Telephony Integration (CTI) was developed to unify the technologies for teleconferencing purposes. The advent of CTI has led to an unprecedented integration of computer and communications technologies, which now enables individuals and businesses to exchange information quickly, efficiently and almost effortlessly.

Growth in CTI closely parallels the dynamic evolution of the Internet. Commercial Internet efforts originally focused on vendors providing basic networking products, as well as service providers offering connectivity and basic Internet services. In the past few years, however, the Internet has become a commodity service with much attention focused on the use of its global information infrastructure to support other commercial services. Today, the Internet functions as a worldwide broadcasting network, a mechanism for information dissemination, and a medium for collaboration and interaction among individuals and their computers without regard for geographic location.

CTI provides computer access and control of telephone functions, as well as telephone access and control of computer functions. CTI also provides a solution to the problem of message management. Users can access and manage their messages

from either the telephone or the PC, no matter where they are physically located. No longer do users have to check three separate places to access their voice mail, facsimiles, and electronic mail.

CTI has existed in commercial form since the mid-1980s, with serious interest developing in this technology during the 1990s. Several factors contributed to the growth of CTI marketplace acceptance, including: definition of international standards for interconnecting telephone and computer systems; industry promotion of mass-marketing application programming interface specifications; improvements in voice processing technologies that provide advanced features and high port densities at attractive prices; offerings by public networks of more and more services which enable computer telephone applications; and, most important, a global economy that is doing business over the telephone at an ever increasing rate.

In a little over a decade, CTI technologies have grown into a multi-billion dollar industry encompassing diverse applications and technologies, ranging from simple voice mail systems to complex multimedia gateways. CTI equipment now includes speech recognition and voice identification hardware, fax servers, and voice response units. The power driving CTI is telephone network access to computer information through such easy-to-use and available terminal devices, as:

- Telephones;
- Analog Display Services Interface phones (ADSI);
- Pagers and Personal Digital Assistants;
- Facsimile machines; and
- Personal Computers.

Businesses leverage the power of CTI systems to improve productivity, provide users with more access to information, and deliver more efficient and cost effective communication services to customers and employees.

Internet Protocol (IP) Telephony is an extension of CTI that enables PC users, via gateways and standard telephony, to make voice telephone calls to anywhere in the world over the Internet or other packet networks, for the price of a local call to an Internet Service Provider. Gateways bring IP Telephony into the mainstream by merging the traditional circuit-switching telephony world with the Internet. Gateways

offer the advantages of IP Telephony to the most common, inexpensive, mobile, and easy to use terminal in the world --the standard telephone.

IP gateways function in the following manner. On one side, the gateway connects to the telephone world via a telephone line plug that enables it to communicate with any telephone in the world. At the other side, the gateway connects to the Internet world, enabling it to communicate with any computer in the world that is connected to the Internet.

The gateway receives a standard telephone signal, digitizes the signal as needed, significantly compresses the signal and packetizes it into IP. It is then routed to its destination over the Internet. A gateway reverses the operation for packets received from the Internet and going out to the telephone. Both operations take place simultaneously, thus allowing for a full duplex (two-way) conversation.

IP Telephony gateways as implemented today usually take users' calls from a PBX, encapsulate the voice information in IP and send it through the company's Wide Area Network (WAN) links to a remote office. The communication signals are transmitted over the Internet using conventional IP formats, such as TCP/IP and SSL, that organize the data into packets for transmission. At the remote office, another gateway extracts the data from the IP packets and sends it to that local PBX, which directs it to the appropriate desktop handset.

IP Telephony will continue to gain in popularity for two reasons. Organizations can receive a significant reduction in long distance costs by using Voice Over Internet Protocol (VoIP) for their long distance calling. Second, because IP uses bandwidth much more efficiently than circuit switching, telephone providers will find that switching over to an IP based network will save enormous costs. Companies such as Net2Phone have demonstrated that it is possible to implement global networks rapidly and much less expensively using IP telephony than with conventional circuit switched hardware. Despite the heavy investment in such hardware by "conventional" telephone companies, there is such a significant economic advantage to VoIP, that even the most entrenched circuit switchers are converting over to a more network-centric view of service provision.

Factors having an adverse effect on commercial acceptability of IP gateway communications, include:

- *Difficult to Operate and Maintain -- At present, video-teleconferencing devices require a great deal of effort and specific knowledge to operate and maintain. Often they require an operator, a network engineer, and a computer technician on each end to assure that the systems are operating properly.*
- *High Cost – IP Telephony devices typically utilize a costly Digital Signal Processor (DSP) that provides complete hardware and software solutions for IP at full-duplex. As the power of DSPs dramatically increases over the next few years, IP Telephony gateways will become mainstream, low-cost, high-density products;*
- *Poor Voice Quality -- IP Telephony has been characterized by poor voice quality, distortions and disruptions in speech, and low reliability. Recently, however, voice quality has begun to improve as a result of technological advances in voice coding, lost packet reconstruction (which makes speech easier to understand), and increased bandwidth capabilities across the Internet;*
- *Prolonged Latency -- Latency, which is an universes in the communication of the sound data, is the primary cause of distortion. Humans can tolerate about 250 milliseconds of latency before it has a noticeable effect, and existing IP products generally exceed this level. Internet Telephony commonly known as "Voice over IP" (VoIP) for example, enables personal computer (PC) users to make voice telephone calls over the Internet or other packet networks via gateways and standard telephones.*

The concept of video phone communication devices appeared in early *Dick Tracy* comic strips in the 1930s, *Flash Gordon* films of the 1940s, *Star Trek* in the 60s, and more recent *Star Wars* movies. The video phone has evolved from a figment of the imagination into a rapidly developing commercial reality, yet most users are dissatisfied with the audio and video resolution that may be obtained from the systems that are commercially available today. Thus far, video phones have been either extensions of the television, or extensions of PCs, none of which really achieve the promise of video telephony that we get a glimpse of in the movies. These systems

derive from a hodgepodge of parts that were never developed, in combination, for the specific purpose of being video phones.

Television-based systems that are produced by such companies as C-Phone of Wilmington, NC and Via TV of London, England, are literally black boxes, each consisting of a microprocessor unit and a tiny camera, that use regular television sets for visual display and conventional telephone lines for transmission. The black box is linked to a television set and connected to the telephone system through a separate cable that plugs into a standard telephone jack. All calls begin as voice-only connections. Each party that is willing and able to appear on the screen pushes a remote control button and, within 30 seconds, a video image appears on each screen.

Commercially available computer-based systems include Microsoft's Net Meeting, CU-SeeMe, 3Com Big Picture Videophones, IRIS Phone, VDO Phone Professional and the Intel Video Phone. These devices convert relatively powerful PCs (i.e. Pentium processors with at least 16 megabytes of memory and appropriate software) equipped with cameras, microphones and other equipment into video phones.

For the following reasons, television-based and computer-based video phone systems have not yet achieved, nor are they likely to achieve, widespread commercial acceptance:

- *Cost and Time Constraints -- Television and computer-based video phones can either operate over the regular telephone network or the Internet. Unlike operating on the Internet, however, utilization of the regular telephone network requires users to pay normal commercial long-distance billing rates. Internet-based systems, on the other hand, pose other problems. These systems are so difficult to use that they cannot be operated by inexperienced persons and even most experienced users must first utilize some other communications means to set up a time for the conference.*
- *Hardware Complexity-- The use of multi-purpose computer hardware and software systems (such as PCs) to perform video phone and conferencing increases the complexity and cost of operation for these devices. This causes a corresponding reduction in reliability and availability. At the same time,*

this approach exposes the video conferencing system to all of the maladies that PCs are subject to, such as computer viruses.

- *Software Complexity -- PC-based software requires extensive expertise to install and to utilize. This creates a barrier for entry into the mass consumer market.*
- *Image and Voice Distortion -- Rapid motion at one end of the line causes image distortion at the other end. If, for example, something moves quickly into and out of the picture, existing video phone devices react by depicting that object as a melange of image blocks. Processes of creating and moving images also place heavy data processing demands on the unit. Because the video image must first be compressed for transmission and then expanded for viewing, the picture and voice get out of phase resulting in what looks like a badly dubbed film.*

The quality of both image and voice on television-based units is clearly better than anything traveling over the Internet, where voices are slightly distorted, and images sometimes dark and blurred. Nearly all video phone systems transmit, at best, only half the number of frames per second that broadcast television uses, making all action look slightly odd. Because of poor image quality, most commercial video systems offer, as an alternative, still images of very high quality produced with snapshot or freeze frame commands.

- *Lack of Interoperability and Industry Standards – Initially, video teleconferencing systems were designed to use ISDN digital lines because they required data rates higher than those available Via common analog telephone lines. Many of those systems are still in use today, even though newer types of data lines can provide much more data throughput at much higher data rates. These are generally the most expensive systems, yet they are trapped by their technology, unable to take advantage of newer forms of data communication. Since these systems could only communicate with other such systems over essentially dedicated ISDN connections, it didn't matter that their protocols were unable to travel through networks. However this inability to route such traffic through networks forced these systems to be the exclusive domain of conference rooms. This inability to deal effectively with the network*

environment and to other devices led to the eventual development of video conferencing standards, collectively as the H.323 standard.

The H.323 standard allows video conferencing between networked systems and provides graceful degradation of service down to speeds of 14,400bps. At the time this standard was developed, modems were achieving those speeds and PCs were beginning to have access to TCP/IP. A number of first generation video phone products were introduced that could connect with each other over the Internet. These devices constitute the majority of installed video conferencing devices in use today. However they are not capable of providing television quality service. That said, as they are the installed base of users, the present invention preferably is compatible with these users

- *Difficulty of Use -- Though there are some very expensive systems in use, they are not easy to use. Generally, they require someone who has been trained in operating these devices. Also, someone who knows how to set up ISDN is required. In addition to tying up these people on each end for every conference, usually a pair of conference rooms is committed to the use of these systems.*

On the low end, the situation is not much better. That is, "low end" video conferencing usually requires a PC, a camera, maybe a video capture card, and software. It usually takes several days or weeks to configure a couple of PCs, set them up on the LAN with IP access, install a camera and get the software to work, install a sound card and microphone, get that software to work, install video conferencing software, and resolve communications issues on both computers. Upon conferencing, each terminus must access the same picture locator server at the agreed upon time, which is a trick in itself. Even after telephone calls to talk each other through establishing the session these sessions often fall miserably. The present invention provides a means for users to avoid all of these problems.

- *Lack of Reliability -- Computers are dynamic devices. Software is installed and subsequently de-installed, operating systems are upgraded, and hardware making up a computer is periodically changed. Since the nature of the computer is dynamic, both reliability and predictability are hindered by its utilization in computer-based phone and conferencing devices. PC based*

systems are susceptible to viruses as well as a variety of user errors. PC based systems generally have hard disks that are the source of catastrophic failures.

- *Lack of Portability -- Current video-conferencing devices require a dedicated space for utilization. Absent is the universal ability to plug these devices into locations where there are compatible network jacks. Even PC based systems are not particularly portable.*

Current technologies have repeatedly failed to provide IP teleconferencing devices that offer acceptable audio and video capabilities and which are also easy to use. No dedicated appliances have been developed for exclusive use in Internet teleconferencing purposes, in part, due to the complexity of adapting high volumes of information for transmission according to Internet protocols.

SUMMARY OF THE INVENTION

The present invention overcomes the problems that are outlined above by providing a dedicated or single-purpose IP teleconferencing appliance that is extremely portable and easy to use in the sense of a device that may be plugged in and turned on for actual teleconferencing use without modification to factory settings and components. For example, the dedicated appliance may provide real-time voice and full motion video at low cost with high reliability, superior fidelity. A unique integration of several components utilizing, for example, wavelet compression codec facilitates these advances without necessarily requiring complicated retrofitting or modifications to an existing PC.

According to the various instrumentalities and embodiments that are described herein a dedicated conferencing system permits a telecommunications conference participant to communicate with another telecommunications conference participant through use of a dedicated device comprising an audio input device, such as a microphone, for use in providing a direct audio input signal. An audio output device, such as a speaker, provides an audio output corresponding to a first compressed audio signal. An audio codec is operably configured for transforming the direct audio input signal into a second compressed audio signal for audio signal transmission purposes and for converting the first compressed audio signal into a form that is usable by the audio output device in providing the audio output. A network communications device

is operably configured for receiving the first compressed audio signal according to an Internet communications protocol and for transmitting the second compressed audio signal according to the internet communications protocol. A controller is programmed with instructions that permit the telecommunications conference participant to communicate with the other telecommunications conference participant through use of the audio input device. The telecommunications conferencing system has essentially no features other than features which are useful for conferencing purposes, and the respective features that are described above are optionally but preferably provided in a single housing that is preconfigured with factory settings.

Further aspects or embodiments of the dedicated teleconferencing system may include a camera for use in producing a first video image signal and a video display device. The network communications device is operably configured for transmitting the first compressed video input signal according to the Internet communications protocol and for receiving a second compressed video signal according to the internet communications protocol. A video codec operably configured for transforming the first video image signal into a first compressed video signal and for translating the second compressed video signal from the other video conference participant into a video output signal that is compatible with use by the video display device. In this case, the program instructions of the controller permit the telecommunications conference participant to communicate with the other telecommunications conference participant through use of the camera and the video display.

The program instructions may comprise instructions for arranging the first compressed video signal and the second compressed audio signal into respective data streams including audio packets and video packets separating the first compressed video signal and the second compressed audio signal for distinct transmission through the network communications device. These program instructions may further be capable of dynamically adjusting a variable packet size of the audio packets based upon sensed errors in receipt of a transmitted signal, such as the first compressed audio signal, the second compressed audio signal, the first compressed video signal and the second compressed video signal. The program instructions of the controller may, in a similar manner, adjust a variable packet size of the video packets based upon sensed errors in receipt of at least one of the first compressed audio signal, the

second compressed audio signal, the first compressed video signal and the second compressed video signal.

Another aspect of the teleconferencing system pertains to program instructions that regulate CPU usage to control the rate of information being transmitted through the network communications device by maintaining a level of CPU utilization below a maximum threshold level. This technique of regulating CPU usage, according to a preferred but optional aspect of the control instructions, optimizes the rate of information transfer by setting the level of CPU utilization just below a rate of utilization that causes an increase in transmission error rates. This functionality may be accomplished, for example, by dynamically adjusting at least one of the audio packet size and the video packet size in response to transmitted error rates.

Additional transmission efficiencies may be realized by inserting a serial identifier into the respective audio packets and video packets to identify a sequential order of packets. This sequential order may, for example, sequentially relate the order of respective audio and video packets in the context of separate audio and video data streams, while also relating the timing of audio packets in relationship to video packets.

Significant testing in the area of audio latency has identified a need to control data packet size. By reducing the packet size, the effects of phase distortion and packet loss are significantly reduced. The tradeoff for the small packet size is network overhead. However, we find that a ¼ to 1/3 increase in network utilization yields a 10:1 improvement in clarity for both audio and video. Accordingly, the program instructions of the controller comprise code for selectively transmitting audio packets in priority preference to video packets, in order to provide an audio latency not greater than 250 ms. This type of latency control may, for example, be accomplished by a feedback loop or by preconfiguring the machine to operate within experimentally established parameters that provide such control.

The concepts described above facilitate extremely high data transmission rates that facilitate a robust multiple teleconferencing capability. Accordingly, a picture-in-picture (PIP) device is optionally but preferably provided for dividing the video display device into respective visual components each allocated to a corresponding conference participant or conference location. A user input device and associated PIP control logic permit the teleconference participant to control the number of respective

visual components on the visual display device. The PIP control logic permits the teleconference participant to scroll through an inventory of teleconference participants when only some of the teleconference participants are represented on the respective visual components at any one time.

A codec is any device or software, such as a dedicated chip with program instructions, that translates incoming and/or outgoing signals. As used herein, the term "codec" pertains to a single device that performs these functions, as well as a logical codec that performs these functions through the use of two or more physical devices. Especially preferred audio and video codecs for use in the teleconferencing system respectively comprise audio and video wavelet compression algorithms.

Additional embodiments and instrumentalities pertain to a method of teleconferencing in which a telecommunications conference participant communicates with another telecommunications conference participant. The method comprising the steps of producing a direct audio input signal, receiving a first compressed audio signal through use of an Internet communications protocol, translating the direct audio input signal through use of an audio codec to compress the direct audio signal and produce a second compressed audio signal, processing the first compressed audio signal through use of an audio codec to transform the first compressed audio signal into a form that is usable by an audio output device in providing an audio output, and transmitting the second compressed audio signal through use of an Internet communications protocol. These steps are performed using a dedicated conferencing system.

The foregoing method pertains to an audio conferencing system that may optionally be expanded to include video processing steps, such as producing a direct video image signal, transforming the direct video image signal into a first compressed video signal through use of a video codec, transmitting the first compressed video input signal according to the Internet communications protocol, receiving a second compressed video signal from the other conference participant, translating the second compressed video signal into a video output signal that is compatible with use by a video display device, and displaying the video output signal through use of the video display device. The foregoing steps are performed using a dedicated conferencing system.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates an exemplary conferencing system in accordance with the concepts described above;

Fig.2 illustrates an exemplary functional diagram of the conferencing system and method;

Fig. 3 is a schematic diagram demonstrating a variety of interconnectivity scenarios;

Fig. 4 is depicts a second embodiment of the conferencing system from a front perspective;

Fig. 5 is depicts a second embodiment of the conferencing system from a rear perspective; and

Figure 6 is a block diagram illustrating a combination of circuits for use in making the conferencing system.

DETAILED DESCRIPTION

There will now be shown and described, by way of example in Fig. 1, a dedicated multimedia conferencing system 100 that permits a telecommunications conference participant to communicate with another telecommunications conference participant through use of a dedicated device. The discussion below teaches by way of example and not by limitation, so the following disclosure should not be construed to unduly limit the scope of the patent claims.

The conferencing system 100 is preferably a dedicated Internet appliance that uses the Internet 102 to serve as a conduit for video phone and conferencing communications. The conferencing system 100 is suited for use by business, government, and academic communities, as well as personal home use. The term "dedicated appliance" is used herein to describe a single purpose telecommunications device having essentially no features that interfere with or are not useable in the context of telecommunications conferencing. The dedicated appliance is preferably constructed to provide a portable plug-in, high resolution video phone and conferencing system that is designed to function without the use of a PC. More particularly, the conferencing system 100 is a real-time, telephonic appliance with audio and visual capabilities. It does not use a personal computer in the traditional sense, and as a result it is not burdened by the additional costs and functionalities

associated with personal computers. For example, in preferred but optional embodiments, system operating overhead is significantly reduced by using an embedded processor that accesses a purpose-built or ROM-stored operating system, as opposed to a commercially available PC operating system having a plethora of unneeded functions with additional associated overhead.

In preferred but optional embodiments the conferencing system 100 is a single-purpose device consisting solely of features that facilitate teleconference communications, thereby ensuring ease of use and reliability. Nevertheless, the conferencing system is also preferably capable of incorporating expansions or enhancements.

The conferencing system 100 may be designed to facilitate only audio, only video, or combined audio visual telecommunications. Video conferencing usually includes both the transmission of video data and audio data. In sending the audio data, an analog voice signal may be captured using built-in stereo microphones 104 and 106, which are preferably high performance low noise microphones having noise canceling circuitry, or an optional single microphone 108 within a plug in telephone handset 110. The analog voice signal is converted to a digital signal that is enveloped in an IP packet and transmitted, for example, on a 10/100 Base-T Ethernet IP Network line 112 according to packetized Internet transmission protocols, Line 112 may also represent high bandwidth cable modem transmissions, and DSL communications. Alternative transmission techniques include, for example, a wireless LAN transmission 114 according to such standards as IEEE 802.11b.

In accordance with the foregoing concepts, the standard analog telephone handset 110 can be used as an audio input device including the microphone 108 for eventual transmission of outgoing audio signals according to a variety of user-selectable transmission techniques. Similarly, the telephone's earpiece 116 including an integral internal speaker (not depicted) can be used as an audio output device, as can broadcast speakers 118 and 120, for presentation of incoming signals. The speakers 118 and 120 are preferably low distortion high fidelity speakers.

One application for the conferencing system 100 is use in transmitting audio signals in voice over Internet Protocol (VoIP). Video over IP may also be added to provide simultaneous video and audio signal transmissions. Alternatively, the Internet 102 may be eliminated and replaced by a direct dial capability linking one

teleconference participant to another. The conferencing system 100 can, for example, function as a PBX/VoIP gateway that takes raw voice data from the telephone handset 110, digitizes the data, encapsulates the data in IP, and directly transmits the data to an identical conferencing system 100 (not depicted), which extracts the data for presentation to a teleconference participant. Thus, a direct dial system may be programmed to utilize IP protocols without ever engaging in an actual transmission over the Internet. In view of the foregoing, the conference system 100 can also be used as a telephone to Internet VoIP gateway.

Whether sending video signals as data or audio signals as data, in either case the conferencing system 100 sends data to another like conferencing system. Usefulness of these portable systems is expanded by providing a variety of optional telecommunications modes, for example, as through the provision of a radio frequency interface 122 that produces the wireless LAN transmission 114 according to such standards as LAN IEEE 802.11 or a satellite IP communications signal. An optical, e.g., infrared, interface 124 may also be utilized as an IP conduit to transmit data. These additional functionalities permit a single dedicated teleconferencing system 100 to connect offices, homes, the people and their devices together across an enterprise using a company's LAN/WAN infrastructure or across the entire world via the Internet 102 where the user may select from among a robust variety of transmission techniques.

Video capture utilities are provided through a full color miniature digital video cameral 126 that captures video images for internal processing and eventual outgoing transmission over the Internet 102. A video display 128, such as a color NTSC display with a touch screen for accepting user input, presents incoming video signals for viewing by the user. The user is optionally but preferably provided with a capability for selecting from the various data transmission modes by interacting with the touch screen functionality of video display 128 or an optional keyboard (not depicted).

The necessary functional components of conferencing system 100 all reside within a housing 130 to provide a compact and portable system. Peripheral devices, such as the telephone handset 110, may optionally be plugged into the housing 130. Telecommunications connections, such as line 112, may be used in combination with

the conferencing system 100 but are not integrally housed with the system components.

The conferencing system 100 allows businesses and government agencies to conduct cost-effective, visually interactive electronic meetings between two or more distant locations. This results in significant reductions in corporate and government travel expenditures while at the same time allowing more individuals to become directly involved in the decision-making process. In the academic community, the conferencing system 100 enables real-time interaction of students and teachers with experts, collaborators, and organizations all over the globe to establish a classroom without a wall.

The functionality of teleconferencing systems like conference system 100 is enhanced when the systems are used through broadband service providers, because improved quality of service is dependent upon additional bandwidth, particularly when simultaneously transmitting audio and visual signals in teleconferencing applications. Lack of data transmission capacity has heretofore proven to be a limiting factor in the use of teleconferencing systems, and preferred service providers are able to offer transmission rates of at least 130 kbps.

Accordingly, the conference system 100 has a variety of features that, in combination, make the most out of the available bandwidth by compressing the audio and/or video signals and packetizing simultaneous transmissions of respective audio and video data streams. Expansion of a packet based network is much less expensive to deploy than a circuit switched one and, consequently, use of packet-based IP telephony technologies is expected to increase substantially over the next few years. As IP telephony enters main stream usage through devices like conferencing system 100, the technology also brings a new suite of advanced capabilities to companies and government agencies, such as group collaboration and video conferencing. Using software that supports these features, people around the world are able to make telephone calls they would never have been able to afford before. Users are able to participate in virtual meetings, enabling them to view conference participants in real time as they speak, and allowing them to collaborate on diagrams (white boarding), in real-time, with people at other locations.

Especially preferred embodiments of conference system 100 advantageously obtain enhanced fidelity by utilizing newly developed wavelet based compression facilities. Recent developments in leading edge wavelet compression techniques have had a positive impact on video technology. Wavelet compression is a process that allows the transmission of electronic images with less demand on bandwidth. It is a highly effective and efficient means of reducing the size of a video stream while maintaining higher quality in the displayed video than is available in other compression techniques. Wavelets have greatly improved the speed with which data can be compressed and are almost 50 times more effective than competing compression methods. As a result of these speed improvements, wavelet technology is proving to be an excellent method for creating live video feeds. As wavelet compression devices become more readily available, their applicability to video communications equipment will increase significantly.

Wavelet compression algorithms are, by way of example, described generally in *McGill University: School of Computer Science Winter 1999 Projects for 308-251B*, DATA STRUCTURES AND ALGORITHMS Project #80: WAVWELET COMPRESSION, which is incorporated herein by reference to the same extent as though fully disclosed herein. The article describes the use of Fast Fourier Transforms in converting signals into Fourier space. While wavelet compression theory is a complex subject, it is sufficient to note that wavelet compression codecs may be purchased on commercial order from suppliers, such as VIANET of Dallas, Texas.

Conventional audio visual transmission technologies typically transmit the audio signal appended as part of the video data stream. It has been discovered that audio latency is affected by the massive video data transmission and, consequently, the conventional practice of appending the audio signal to the video signal fails to provide flexibility in controlling the total amount of data in a manner that permits control of audio latency and increasing amounts of data transmission errors that arise through CPU utilization generally in excess of about 80% of CPU capacity. Accordingly, it is preferred that the audio and video data signals are broken out into respective packetized streams for separate simultaneous broadband transmission. As described in more detail below, the packets may be sized and transmitted in a

selective manner that maintains audio latency within acceptable parameters while minimizing transmission errors that arise through excessive CPU utilization rates.

The video data stream is processed to reduce the ratio of transmission of video packets while giving audio packets a higher transmission priority, e.g., three audio packets to two video packets, in order to preserve CPU utilization under 80% while maintaining less than 250 ms audio latency.

IP transmission techniques typically involve the organization of data into packets that are transmitted from a sending node to a destination node. The destination node performs a data verification calculation, e.g., a checksum, and compares the result to the result of an identical calculation that is performed by the sending node prior to data transmission. If the results match, then the packet is deemed to be error free and is accepted by the destination node. If the results match, then the packet is usually discarded and the destination node sends a handshake signal to the sending node to trigger resending of the packet. These principles create a situation where error-free transmission rates are maximized by sending larger packet sizes, but if error intensive communications are realized, overall data transmission rates may increase by sending smaller packet sizes due to the avoidance of retransmitting larger packet sizes. Accordingly, it is a preferred feature of conferencing system 100 to dynamically adjust the packet sizes during a teleconference depending upon the number of transmission errors that are experienced. The packet sizes are preferably adjusted on the basis of an algorithm that tracks the number of errors and assigns a packet size based upon a correlation of empirical results. Packet sizes that are adjusted according to these principles typically, but not necessarily, fall within the range from one to three kb.

As an alternative to wavelet compression, the conference system 100 can establish and maintain video conferences with other units that adhere to the H.323 body of compression standards, which the user may selectively access by interacting with menu options presented on the touch screen display 128. Lack of both interoperability and industry standards has been largely reduced by the introduction of H.323 and H.324 industry standards from the International Telecommunications Union (ITU). H.323 provides a foundation for interoperability and high quality video and audio data over telephone lines. H.324 specifies a common method for video, voice and data to be shared simultaneously over high speed dial-up modem

connections. These standards are presently incompatible with wavelet compression techniques. The conference system 100 is preferably also capable of using H.261/H.263, compression methods, as required, and functionality is preferably provided such that a standard telephone can send and receive Voice over IP communications in accordance with the Belcore Specifications.

The video display 128 has a preferred but optional white boarding capability that permits multiple users to interact with each other by simply writing or drawing on an area of their local display screens with their finger or a stylus. In addition, the conferencing system 100 has a preferred but optional capability to interface with existing encryption equipment to provide users with reasonable levels of security.

In addition to consumer and business applications described thus far, there are a host of other potential users of the conferencing system 100. For example, some U.S. Department of Defense (DOD) agencies have commented that the conferencing system 100 is an invaluable military tool for directly or indirectly enhancing "battlefield awareness" through more effective information acquisition, precision information direction at the working personnel level, and consistent battle space understanding. The conferencing system 100 provides battlefield commanders with timely, high quality, encryptable information, including surveillance reports, target designations, and battle damage assessments, to elevate the level and speed of military leaders' cognitive understanding of battlefield dynamics, through multi-user (Joint Commands) dissemination and integration of different media products (e.g., voice, maps, and photographs).

A translation server 130 may be accessed through use of the Internet 102, such that VoIP transmissions are submitted for conventional voice recognition that converts speech to text, translates the text to another language text, e.g., from English to Spanish, and converts the translated text into speech using conventional speech generation software. After this processing, the signal is transmitted through the Internet to a destination address. The translation server 130 is specially adapted for use with conferencing system 100 and like systems because translation server 130 is provided with audio and video codecs and related circuitry for processing audio and video images to provide translated speech in patterns that are compatible with conferencing system 100 data and addressing formats.

Figure 2 is a block schematic diagram of logical components that may be assembled to form the conferencing system 100. In Fig. 2, like numbering of identical components has been retained with respect to Fig. 1, except that suffixes A, B, C, and D have been added to represent physical structures that may be programmed for use as different logical components. More specifically, the physical structures shown in Fig. 1 may have a plurality of logical functions pursuant to the discussion above, depending upon the programmed state of conferencing system 100. Thus, for example, line 112 shown in Fig. 1 may represent input components 112A and 112B depending upon whether circuitry internal to the teleconferencing system 100 is programmed or built to function as an Ethernet interface 112A or a conventional AC LAN 112B. Similarly, the line 112 may represent output components 112 C and 112 D depending upon whether circuitry internal to the teleconferencing system 100 is programmed or built to function as an Ethernet interface 112A or a conventional AC LAN 112B. In like manner, the radio frequency device 122 may function as wireless LAN input device 122A or a wireless LAN output device 122B, just as the optical interface 124 may be used as an optical input device 124A and an optical output device 124B. The display 128 has two logical functions including use as a touch screen input device 128A and a display output device 128B.

The heart of teleconferencing system 100 is a controller 200. The controller 200 is programmed with operating instructions that provide the functionalities which are described above. Physical structures for implementing the logical controller 200 shown in Fig. 2 may include, for example, a central processor connected to EEPROMS on which a single purpose operating system is stored together with firmware, as well as distributed processing environments where multiple processors are assigned different functions or complimentary assisted functions. The controller provides any necessary processing and control that is required for accepting inputs and converting the inputs to outputs for teleconferencing purposes. The program instructions may be provided using any manner of program instructions that are compatible with the selected hardware implementation.

A data storage device 202 may include magnetic data storage, optical data storage, or storage in nonvolatile memory. The data storage device preferably includes a removable data storage medium, such as an optical disk or CD-ROM, so that business or technical data, as well as selected audio and video data, may be

retained as specified by the user. For example, a battlefield commander may capture a noise or image for subsequent dissemination and military analysis. Alternatively, participants in a business teleconference may capture the contents of their jointly developed whiteboard and store the same for future use, or a team-built document or spreadsheet can be stored and recalled in an identical manner.

A video wavelet codec 204A accepts a digital video image signal from camera 126 and transforms the signal through use of a wavelet compression algorithm. An intermediate analog to digital converter (not shown) may be positioned between the camera 126 and the video wavelet codec 204A to convert the analog video image signal into a digital signal. The signal from wavelet video codec 204A is transmitted to a network communications output device, such as the IR Interface 124B, the Ethernet interface 112C, the AC LAN 112D or the wireless LAN 122B, as directed by controller 200 pursuant to user specified parameters selecting the mode of output through menu-driven interaction with the touch screen 128A. The signal from camera 126 is alternatively processed and compressed by the H.323/324 video processor 206A according to user-specified parameters for eventual output on the video output devices. Controller 200 forms into data packets the video data stream from either the video wavelet codec 204A or the H.323/324 video processor 206A and assigns serial numbers to the data packets placing the individual video data packets in sequential order. A data header preferably identifies the individual packets as video data packets.

An audio input processor 208A accepts input from the audio input devices including microphones 104, 106 or the telephone handset microphone 108. The audio input processor 208A is preferably a wavelet compression codec that is specifically designed for audio applications. The audio input processor 208A preferably includes an analog to digital converter that converts the analog audio signal to a digital signal prior to submitting the signal for wavelet compression processing. Controller 200 forms into data packets the audio data stream from audio input processor 208A and assigns serial numbers to the data packets placing the individual audio data packets in sequential order.

A data header preferably identifies the individual packets as audio data packets. The sequential ordering of audio and video data packets preferably intermixes the order of audio and video data packets, for example, such that a first

audio packet is assigned a serial number of one, a first video data packet is assigned a two, a second video data packet is assigned a three and a second audio data packet is assigned a four. The relative ordering of audio and video data packets permits the data stream, upon receipt by an identical teleconferencing system 100, to process the data packets in a manner that plays back the transmitted signals in an order that sequentially assigns audio data packets to video data packets for simultaneous playback. For example, the audio packets represented by serial numbers one and four may be processed and played back to stretch over the video interval represented by packet serial numbers two and three.

Controller 200 receives audio and/or video inputs that are transmitted through network communications input devices including the optical interface 122A, the Ethernet interface 112A, the AC LAN interface 112B, or the wireless LAN interface 122A. These signals arrive in respective data packets that controller 200 processes as described above for synchronized playback. The sequentially combined video data packets are, pursuant to menu driven user specifications by interaction with touch screen 128A, submitted to either a video wavelet codec 204B H.323/324 processor 206B for decompression by inverse transformation and output to the display 128B. The audio data packets are similarly combined in sequential order and submitted to an audio codec 208B for output as an analog signal to either speakers 118, 120 or the telephone speaker 116, according to user specifications.

Controller 200 may also be provided with encryption/de-encryption faculties, or dedicated circuitry may be provided for this purpose.

Figure 3 shows use of the teleconferencing system 100 in a possible video-conference configuration 300 that includes a plurality of identical systems 300 and 302, as well as a conventional teleconferencing system on a traditional PC 304, all of which are mutually participating in a teleconference. Any number of conferencing systems may be connected to and disconnected from the configuration 300 during the course of the teleconference, and the individual teleconference systems will dynamically adjust to accommodate the differing number of users. Conferencing systems 302 and 304 are connected to a 10/100 Base T network, the respective components of which are identified as 306 and 308. Conferencing system 100 is connected to both a 10/100BaseT network 308 and an AC power line network 312. Conferencing system 300 is only connected to the AC Power line network 312.

Teleconferencing systems 100 and 302 are also connected to a local area network (LAN) or high speed Internet connection 102.

The configuration in Figure 3 is used to illustrate the various connection methods of the video-conferencing devices. Some examples of possible conferencing scenarios are provided below.

Teleconferencing system 100 places a call to teleconferencing system 302 over the 10/100 Base T network 310-306. Wavelet compression is used to achieve, for example, 30fps full screen video. The connection may be via local network or Internet 102.

Teleconferencing system 100 places a call to teleconferencing system 304 over the 10/100BaseT network 310-308. In this case, H.261 or H.263 is used to remain compatible with other manufacturers of video conferencing equipment. The video quality is subject to the limits of the H.261 or H.263 standards because teleconferencing system 100 senses these protocols in transmissions from teleconferencing system 304. The connection may be via local network or Internet 102.

Teleconferencing system 100 places a call to teleconferencing system 300 over the AC Power Line network 312. Wavelet compression is used to achieve 30fps full screen video.

The above examples show teleconferencing system 100 placing all of the calls, however, the calls may be initiated by any of the teleconferencing systems 100, 300, 302, or 304.

Fig. 4 depicts a schematic front view of a second embodiment, namely, conferencing system 400 from a front perspective view. Conferencing system 400 preferably comprises a 10.4" Color NTSC LCD display 402 that includes an integral built-in touch-screen, a centrally disposed built-in color NTSC camera 404, centrally disposed stereo microphones 406 and 408, and embedded circuitry for video and audio processing (not shown). Internal to the conferencing system 400 are various cables, connectors and circuit boards containing the necessary circuitry to process the video and audio

Fig. 5 depicts a rear view of conferencing system 400 that reveals stereo speakers 500 and 502, as well as video In/Out connectors 504, audio left channel In/Out connectors 506, audio right channel In/Out connectors 508, a mouse connector

510, a keyboard connector 512, a RJ45 network connector 514 for 10/100BaseT Ethernet, and an AC power connector 516, all mounted on a connector panel 518. A single housing 520 contains all of these components, as well as the internal circuitry that provides conferencing capabilities through use of these components in an identical manner with respect to the teleconferencing system 100 that is described above.

The color NTSC display 402 is used to display the remote video with a picture-in-picture feature showing the local video being sent. The display 402 also displays menus and messages for interactive user setup and configuration. Touch-screen functionality is integrated into the display 402 and allows the user to operate the teleconferencing system 400 without requiring the use of keyboard or mouse.

The color NTSC camera 404 is embedded into the teleconferencing system 400 and is the source of the local video signal that is processed and transmitted to a remote location. This local video signal, as it is being sent, is shown on the display 402 in a picture-in-picture format along with images of other teleconference participants.

The sound system consists of the built-in stereo microphones 406 and 408, as well as stereo speakers 500 and 502. Also available on the connector panel 518 are separate audio/video in and out connectors 504, 506, 508 for connecting external audio and video sources, displays and sound systems.

Optional connectors found on the connector panel 518 include a mouse connector 510 and a keyboard connector 512. These connectors allow the user to utilize a mouse and keyboard instead of the touch-screen functionality of display 402 in instances where the touch-screen functionality is impractical or undesirable.

The RJ45 network connector 510 provides the connection to a 10/100BaseT network card. This connection can be used to connect to a switch, server or other network devices including another teleconferencing system 400.

The AC power connector 516 provides the power for the conferencing system 400 using standard 100 volt alternating current, or another standard depending upon locale, and also provides and alternative network connection to other teleconferencing systems using a building's power lines.

Upon connecting power to the conferencing system 400, the display 402 presents the local video in a small portion of the display 402. The user may tap the

upper left of the display 402 to bring up a menu with various setup options. The first time the conferencing system 400 is used, the user is preferably prompted to enter an IP address, unless one is provided by a DHCP server. Other setup options may include connection speed, picture-in-picture size and various esthetic settings.

The user may initiate a call by either interacting with the touch-screen display to select a person from a phone-book database or by dialing the number of the remote device using a dial-pad on the touch-screen display 402. If the remote device is another device having identical capabilities to teleconferencing systems 100 and 400, then a device handshake assures that the superior wavelet compression method is used for video compression, providing high resolution, full screen color video. If the remote device is something other than a wavelet compression compatible system, for example, a PC based system running NetMeeting, then either H.261 or H.263 protocol will be used for compatibility purposes. In this case, the video compression and quality is limited to the constraints in those standards.

As previously mentioned, the network connection can be either 10/100BaseT or AC power line. This selection can be made in the setup menu when booting conference system 400. Network connections can be LAN, WAN or Internet based, provided a high-speed connection is being used.

The conferencing systems 100 and 400 are extremely easy to operate. All that is required after initial setup is for the systems to be plugged into an AC power outlet and they can, for example, communicate with any other compatible device on the AC power line network. Additionally, a RJ45 10/100BaseT network cable can be connected to allow the systems to communicate with any video-conferencing system on the network or over the Internet. The user interface in its simplest form only requires that a phone number be entered, or that one is selected from an address or phone-book database.

Fig. 6 depicts a functional block diagram for exemplary circuitry 600 inside the conferencing systems 100 or 400. A description of the signal processing follows.

The video signal is generated using a NTSC camera 602, preferably having at least 320 lines of resolution. The camera 602 provides NTSC video to be digitized, compressed and sent across the network to a remote video-teleconferencing device. Camera 602 is connected to a video decoder chip 603, which separates the NTSC signal for transmission to both a wavelet codec 604 and a processor 606 that is

programmed to provide instructions causing the operations attributed to controller 200 (see Fig. 2). The video decoder chip 603, e.g., a SA711A circuit, is responsible for taking the NTSC video from the local camera and converting it to YUV(CCIR656) so that the wavelet codec 604 and the processor 606 can process the video data.

The processor 606 is preferably an embedded processor having an exclusive telecommunications processing function. For example, the PTM1300EBEA processors that may be purchased from Trimedia Technologies of Austin, Texas are intended for video, audio and graphics purposes. These chips operate at speeds exceeding 166 Mhz and are capable of 6.5 billion operations per second. Accordingly, commercial varieties of processor 606 have ample power to compress and decompress many video and audio formats and are well suited for video-conferencing applications.

The processor 606 accesses SDRAM 606A for memory, EEPROM 606B for boot strapping code and another EEPROM 606 for the program code. The SDRAM 606A, e.g., a HM5264165FTT chip is used to store temporary data during compression algorithms. The boot EEPROM 606B, e.g., an AT24C16 chip, stores the first few instructions for the processor 606, initiates a basic setup, and points to the EEPROM 606C containing the program code. The EEPROM 606C, e.g., a AT27C040 chip, stores the program code for the processor 606. Processor 606 communicates primarily with seven other devices to provide data transfer and control instructions.

The wavelet codec 604, e.g., an AD601LCJST compression chip, allows for very high quality, full size video to be sent at fairly high compression rates. Wavelet codec 604 requires DRAM 605 to operate. For example., DRAM in the form of a HM514265 circuit, is accessed by the Wavelet codec 604 to temporarily store data during compression.

The best video quality and compression can be obtained by using the wavelet codec 604 between wavelet compression-compatible systems, but not all systems are wavelet compression-compatible. For interfacing to non-wavelet compatible systems, the processor 606 can convert the video into common standards like H.261 and H.263. The wavelet codec 604 compresses the video with assistance from a digital signal processor (DSP) 608 and feeds the resultant signal into the processor 606. The DSP 608, e.g., an ADSP-2185 circuit, is used for computing the Bin Width calculations for

the wavelet codecs 604 and 610 to accomplish both compression and decompression of data. DSP 608 also is the data interface between both codecs 608 610 and the processor (36). DSP 608 requires SRAM 608A and an EPROM 608B to run. The SRAM 608A, e.g., a HM621664 chip, is used by the DSP 608 during Bin Width Calculations. The EEPROM 608B, e.g., a AT27C040 chip, stores code for operating DSP 608B.

The processor 606 selects the appropriate video input based on the connection type with the other system. If the other system is wavelet compression compatible, processor 606 selects the compression de-compression pathway including wavelet codecs 606 and 610. If the other system is non-wavelet compression compatible, e.g., NetMeeting, the processor 606, if of a TriMedia variety, uses its own internal H.261 and H263 algorithms for video compression to remain compliant with these conventional standards.

The incoming audio signal is generated by a built in microphone 612 and an automatic gain control (AGC) amp 614. The AGC amp 614, e.g., a SSm2166P chip, receives an audio signal from the microphone and provides a constant output to an audio codec 620, which thereby receives an incoming audio signal that is smooth and at a consistent level, which is desirable during video and audio conferences. The microphone 612 is also built-in and is responsible for capturing the local audio. The audio signal is fed into an automatic gain control (AGC) circuit 614. The audio output signal is heard through use of built-in speakers 616, which are driven by an audio amplifier 618 which may be specified, for example, as an LM1877N chip. The speakers 616 are built into the system and present the far end audio to the intended recipients, preferably in stereo format.

Both the speakers 616 and the microphone 612 are interfaced to the audio codec 620, which converts the audio from analog to digital and vice-versa while appropriately compressing and de-compressing the audio signal, preferably using audio-specific wavelet compression algorithms. The audio codec 620, e.g., a UDA1344TS chip, is responsible for communicating digital audio signals to and from the processor 606. The audio codec 620 receives the incoming audio signal from the AGC amplifier 614, which is connected to the microphone 612, and sends outgoing audio signals to the audio amplifier 618, which drives the speakers 616.

The processor 606 provides control instructions for packetizing the respective data streams with serial numbers as described above, combines the video packets with the digital audio packets and transfers the audio and video signals to an Ethernet MAC chip 622, which is formats according to any packetized internet transmission protocol. The Ethernet MAC chip 622 sends the packetized data to an Ethernet PHY (physical layer driver) chip 624 and a power line interface chip 626. Thus, circuitry 600 can communicate over 10/100 Base T using an RJ45 connector 626 and also over common AC power lines 628. A power supply 630 is used to convert AC line voltage (110vac, 60 Hz) to various DC voltages that are required by the circuitry.

The Ethernet Mac chip 622, e.g., a LAN91C100 chip, provides the processor 606 with network capability. The Ethernet Mac chip 622 takes data from the processor 606 and creates Ethernet packets to be sent via CAT5 through the RJ45 connector 626 or AC power lines 628. For 10/100BaseT operation, the data is sent to the Ethernet PHY chip 624. For power line data transmission, the data is sent to the power line interface 627. The operations of Ethernet MAC chip 622 require a small amount of SRAM 622A, e.g., an IS61C3216 chip.

The Ethernet PHY chip 624, e.g., a LAN83C180 chip, is the physical interface for the 10/100BaseT network that is accessed though RJ45 connector 626. The Ethernet PHY chip 624 receives data from the Ethernet MAC chip 622 and converts the data into the voltages that are necessary for 10/100BaseT communications over CAT5 cable.

The power line interface 627 allows 10Mbit/s network communication over common AC power line (20), and interfaces with the Ethernet MAC chip 622.

A combined packetized digital video and audio data stream is received in 10/100 Base T format either over CAT5 cable through the RJ45 connector 626 or via AC the power line connection 628. The incoming data passes onto the processor 606. If the information is from a wavelet compression-compatible device, the processor 606 passes this video data onto the wavelet codec chip 610 for decompression. DRAM 610A, e.g., a HM514265 chip, is used by the wavelet codec 610 to temporarily store data during data decompression operations. The de-compressed data then transfers to a video encoder chip 632. The video encoder chip 632, e.g., an ADV7175A circuit, receives video in YUV(CCIR656) format and converts it into NTSC. This NTSC video is sent to the PIP chip 634. The video encoder chip 632

receives its video either from the wavelet decompression codec 610 or directly from the processor 606 depending on which video compression method is being used.

If the incoming data stream is non-wavelet compression compatible, the processor 606, if of a TriMedia variety, uses internal H.261 or H.263 algorithms to decompress the video internally and directly sends the de-compressed data to the video encoder chip 632, which converts all incoming video data from YUV (CCIR656) to NTSC and sends the data to a picture-in-picture chip 634, which superimposes the video image into a corresponding area of the display that is allocated to the logical location generating the image. From here, the video signal travels through a video overlay chip 636.

The video overlay chip 636, e.g., a UPD6465GT chip, receives instructions from the processor 606 and overlays text menus on the video. This faculty provides the processor 606 with a way of displaying menus and information on the NTSC display 638 and also responds to touch-screen circuitry 640. The video overlay chip 636 receives a combined video image from the PIP chip 634, adds the appropriate text, and sends the composite image to the NTSC LCD display 638.

The PIP chip 634, e.g., a SDA9288XGEG chip, combines the video images from the far end conference participants with the video coming from the local camera 602. This faculty enables the user to see the video he or she is sending out in a small corner of the display. The PIP chip 634 receives the far side NTSC video from the video encoder chip 632 and the local camera video directly from the camera 602. The combined video image is sent onto the video overlay chip 636.

The processor 606 directs the implementation of menu-driven user-specified options, such as where user-specified menu instructions may control the nature of the PIP image, for example, to limit the number of participant images that are simultaneously displayed at any one time or to scroll through a plurality of participant images.

The combined video images eventually reach a 10.4" NTSC LCD display 638, which has integral touch screen circuitry for use in accepting user commands. The color display 638 is, for example, a 10.4" Color Flat Screen NTSC LCD with built in touch-screen circuitry 648, and it provides the user with the far end video, local video and various menu screens and messages. The touch-screen circuitry 648 provides a serial output to the processor 606 based on which part of the screen was touched, and

comes pre-integrated with the NTSC Color Display 638. The touch-screen circuitry 648 provides the users with the ability to quickly select functions and options without necessarily using a mouse or keyboard.

The processor606 is coupled with sufficient EEPROM memory storage 640 to boot the TriMedia processor with a dedicated operating system that is provided by the manufacturer. Program instructions for accomplishing the foregoing functions are similarly stored in EEPROM 642, and SDRAM 644 is sufficient to facilitate operations of the TriMedia processor 606.

The following discussion provides specific examples of commercially available components that may be combined to assembly the circuit shown in Fig. 6.

Many variations of the illustrated example can be deployed. For example, the wavelet codec chipsets 604 and 610, together with associated memory 604 A and 610A could be incorporated on a separate board that interfaces with a main (or mother) board. Further the DSP chip 608 could be replaced by a microprocessor and appropriate software stored in, for example, flash memory. More generally, the functional components described in the context of Fig. 6 can be combined or separated into a variety of different hardware components.

Finally, in a preferred embodiment of the present invention, to the extent an operation system is necessary, Linux is preferably employed. Of course, other operating systems may be implemented depending on circumstances and designer preferences.

Therefore, the invention in its broader aspects is not limited to the specific details, representative devices and methods, and illustrative examples shown and described. Accordingly, departures may be made from such details without departing from the spirit or scope of the general inventive concept as defined by the appended claims and their equivalents.

CLAIMS

We claim:

1. A dedicated conferencing system that permits a telecommunications conference participant to communicate with another telecommunications conference participant, comprising:

an audio input device for use in providing a direct audio input signal;

an audio output device for use in providing an audio output corresponding to a first compressed audio signal;

an audio codec operably configured for transforming the direct audio input signal into a second compressed audio signal for audio signal transmission purposes and for converting the first compressed audio signal into a form that is usable by the audio output device in providing the audio output;

a network communications device operably configured for receiving the first compressed audio signal according to an internet communications protocol and for transmitting the second compressed audio signal according to the internet communications protocol; and

a controller programmed with program instructions that permit the telecommunications conference participant to communicate with the other telecommunications conference participant through use of the audio input device,

the telecommunications conferencing system having essentially no features other than features which are useful for conferencing purposes.

2. The conferencing system as set forth in claim 1, further comprising:

a camera for use in producing a first video image signal;

a video display device,

the network communications device being operably configured for transmitting the first compressed video input signal according to the internet communications protocol and for receiving a second compressed video signal according to the internet communications protocol, and

a video codec operably configured for transforming the first video image signal into a first compressed video signal and for translating the second compressed video signal from the other video conference participant into a video output signal that is compatible with use by the video display device,

the program instructions of the controller permitting the telecommunications conference participant to communicate with the other telecommunications conference participant through use of the camera and the video display.

3. The conferencing system as set forth in claim 2, wherein the program instructions of the controller comprise program instructions for arranging the first compressed video signal and the second compressed audio signal into respective data streams including audio packets and video packets separating the first compressed video signal and the second compressed audio signal for distinct transmission through the network communications device.

4. The conferencing system as set forth in claim 3, wherein the program instructions of the controller comprise means for dynamically adjusting a variable packet size of the audio packets based upon sensed errors in receipt of at least one of the first compressed audio signal, the second compressed audio signal, the first compressed video signal and the second compressed video signal.

5. The conferencing system as set forth in claim 4, wherein the program instructions of the controller comprise means for adjusting a variable packet size of the video packets based upon sensed errors in receipt of at least one of the first compressed audio signal, the second compressed audio signal, the first compressed video signal and the second compressed video signal.

6. The conferencing system as set forth in claim 5, wherein the program instructions of the controller comprise means for regulating CPU usage to control the rate of information that is transmitted through the network communications device by maintaining a level of CPU utilization below a maximum threshold level.

7. The conferencing system as set forth in claim 6, wherein the means for regulating includes means for optimizing the rate of information transfer by setting the level of CPU utilization just below a rate of utilization that causes an increase in transmission error rates.

8. The conferencing system as set forth in claim 7, wherein the means for optimizing includes means for adjusting at least one of the audio packet size and the video packet size.

9. The conferencing system as set forth in claim 5, wherein the means for adjusting includes means for adjusting at least one of the audio packet size and the video packet size.

10. The conferencing system as set forth in claim 3, wherein the program instructions for the controller comprise means for inserting a serial identifier into the respective audio packets and video packets to identify a sequential order of packets.

11. The conferencing system as set forth in claim 3, wherein the program instructions of the controller comprise means for selectively transmitting audio packets in priority preference to video packets, in order to provide an audio latency not greater than 250 ms.

12. The conferencing system as set forth in claim 2, comprising a picture-in-picture device for dividing the video display device into respective visual components each allocated to a corresponding conference participant.

13. The conferencing system as set forth in claim 12, comprising a user input device and associated PIP control logic permitting the teleconference participant to control the number of respective visual components on the visual display device.

14. The conferencing system as set forth in claim 13, wherein the PIP control logic permits the teleconference participant to scroll through an inventory of teleconference participants when only some of the teleconference participants being represented on the respective visual components at any one time.

15. The conferencing system as set forth in claim 12, wherein the video codec comprises a video wavelet compression algorithm.

16. The conferencing system as set forth in claim 1, wherein the audio codec comprises an audio wavelet compression algorithm.

17. The conferencing system as set forth in claim 1, wherein the program instructions of the controller comprise program instructions for arranging the second compressed audio signal into an audio packet for distinct transmission through the network communications device and the program instructions of the controller comprise means for dynamically adjusting a variable packet size of the audio packets based upon sensed errors in receipt of at least one of the first compressed audio signal and the second compressed audio signal.

18. The conferencing system as set forth in claim 17, wherein the program instructions of the controller comprise means for regulating CPU usage to control the rate of information that is transmitted through the network communications device by maintaining a level of CPU utilization below a maximum threshold level.

19. The conferencing system as set forth in claim 18, wherein the means for regulating includes means for optimizing the rate of information transfer by setting the level of CPU utilization just below a rate of utilization that causes an increase in transmission error rates.

20. The conferencing system as set forth in claim 19, wherein the means for optimizing includes means for adjusting the audio packet size.

21. The conferencing system as set forth in claim 17, wherein the means for dynamically adjusting includes means for adjusting at least one of the audio packet size and the video packet size.

22. The conferencing system as set forth in claim 1, wherein the program instructions for the controller comprise means for inserting a serial identifier into the respective audio packets and video packets to identify a sequential order of packets.

23. The conferencing system as set forth in claim 1, wherein the program instructions of the controller comprise means for selectively transmitting audio packets in priority preference to video packets, in order to provide an audio latency not greater than 250 ms.

24. A method of teleconferencing in which a telecommunications conference participant communicates with another telecommunications conference participant, the method comprising the steps of:

a.) producing a direct audio input signal;

b.) receiving a first compressed audio signal through use of an internet communications protocol;

c.) translating the direct audio input signal through use of an audio codec to compress the direct audio signal and produce a second compressed audio signal;

d.) processing the first compressed audio signal through use of an audio codec to transform the first compressed audio signal into a form that is usable by an audio output device in providing an audio output; and

e.) transmitting the second compressed audio signal through use of an internet communications protocol,

wherein the foregoing steps a.) through e.) are performed using a dedicated conferencing system.

25. The method according to claim 24, further comprising the steps of

f.) producing a direct video image signal;

g.) transforming the direct video image signal into a first compressed video signal through use of a video codec ; and

h.) transmitting the first compressed video input signal according to the internet communications protocol;

i.) receiving a second compressed video signal from the other conference participant;

j.) translating the second compressed video signal into a video output signal that is compatible with use by a video display device; and

k.) displaying the video output signal through use of the video display device,

wherein the foregoing steps f.) through k.) are performed using a dedicated conferencing system.

26. The method according to claim 25, wherein the transmitting steps e.) and h.) comprise arranging the first compressed video signal and the second compressed audio signal into respective data streams including audio packets and video packets separating the first compressed video signal and the second compressed audio signal for distinct transmission through the network communications device.

27. The method according to claim 26, wherein the transmitting step e.) comprises dynamically adjusting a variable packet size of the audio packets based upon sensed errors in receipt of at least one of the first compressed audio signal, the second compressed audio signal, the first compressed video signal and the second compressed video signal.

28. The method according to claim 27, wherein the transmitting step h.) comprises dynamically adjusting a variable packet size of the video packets based upon sensed errors in receipt of at least one of the first compressed audio signal, the second compressed audio signal, the first compressed video signal and the second compressed video signal.

29. The method according to claim 27, comprising a step of regulating CPU usage to control the rate of information that is transmitted through the network communications device by maintaining a level of CPU utilization below a maximum threshold level.

30. The method according to claim 29, wherein the step of regulating comprises optimizing the rate of information transfer by setting the level of CPU utilization just below a rate of utilization that causes an increase in transmission error rates.

31. The method according to claim 30, wherein the step of optimizing includes adjusting at least one of the audio packet size and the video packet size.

32. The method according to claim 28, wherein the dynamically adjusting step includes adjusting at least one of the audio packet size and the video packet size.

33. The method according to claim 27, wherein transmitting steps e.) and f.) comprise inserting a serial identifier into the respective audio packets and video packets to identify a sequential order of packets.

34. The method according to claim 27, wherein the transmitting steps e.) and f.) comprise selectively transmitting audio packets in priority preference to video packets, in order to provide an audio latency not greater than 250 ms.

35. The method according to claim 26, wherein the displaying step comprises comprising displaying a plurality of video images on the video display device through use of a picture-in-picture device that divides the video display device into respective visual components each allocated to a corresponding conference participant.

36. The method according to claim 35, comprising permitting the teleconference participant to control the number of respective visual components on the visual display device.

37. The method according to claim 36, wherein the permitting step comprises scrolling through an inventory of teleconference participants when only some of the teleconference participants being represented on the respective visual components at any one time.

38. The method according to claim 26, wherein the video codec comprises a video wavelet compression algorithm that is used ion the transforming step c.).

39. The method according to claim 25, wherein the audio codec comprises an audio wavelet compression algorithm that is used ion the transforming step g.)..

40. The method according to claim 25 comprising a step of teleconferencing with at least ten teleconference participants in different locations with each participant experiencing less than 250 ms in audio latency.

41. The method according to claim 26, wherein the transmitting steps e.) and h.) comprise selectively transmitting audio packets in priority preference to video packets, in order to provide an audio latency not greater than 250 ms.

42. The method according to claim 25, wherein the displaying step k.) comprises comprising displaying a plurality of video images on the video display device through use of a picture-in-picture device that divides the video display device into respective visual components each allocated to a corresponding conference participant.

43. The method according to claim 42, comprising permitting the teleconference participant to control the number of respective visual components on the visual display device.

44. The method according to claim 43, wherein the permitting step comprises scrolling through an inventory of teleconference participants when only some of the teleconference participants being represented on the respective visual components at any one time.

45. The method according to claim 25, wherein the video codec comprises a video wavelet compression algorithm that is used ion the transforming step c.).

46. The method according to claim 24, wherein the audio codec comprises an audio wavelet compression algorithm that is used ion the transforming step g.).

47. The method according to claim 24 comprising a step of teleconferencing with at least ten teleconference participants in different locations with each participant experiencing less than 250 ms in audio latency.

48. The conferencing system as set forth in claim 24, wherein the program instructions of the controller comprise program instructions for arranging the second compressed audio signal into audio packets for distinct transmission through the network communications device.

49. The method according to claim 24, wherein the transmitting steps e.) and h.) comprise arranging the second compressed audio signal into audio packets for distinct transmission through the network communications device.

50. The method according to claim 49, wherein the transmitting step e.) comprises dynamically adjusting a variable packet size of the audio packets based

upon sensed errors in receipt of at least one of the first compressed audio signal and the second compressed audio signal.

51. The method according to claim 50, comprising a step of regulating CPU usage to control the rate of information that is transmitted through the network communications device by maintaining a level of CPU utilization below a maximum threshold level.

52. The method according to claim 29, wherein the step of regulating comprises optimizing the rate of information transfer by setting the level of CPU utilization just below a rate of utilization that causes an increase in transmission error rates.

53. The method according to claim 52, wherein the step of optimizing includes adjusting the audio packet size.

54. The method according to claim 24, wherein the transmitting step e.) comprises transmitting the second compressed audio signal to a translation server for translation of a spoken language.

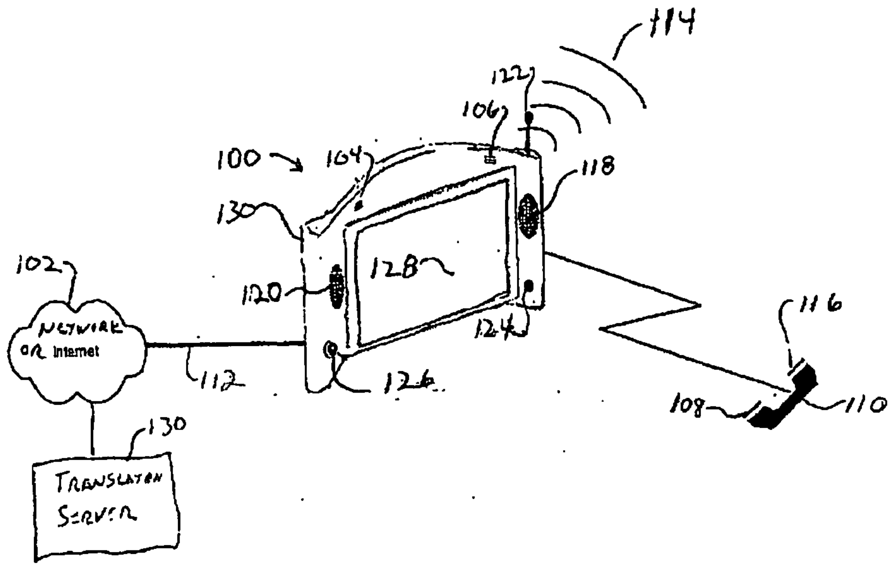

114
122
106
100
104
118
130
102
128
120
NETWORK OR Internet
124
116
112
126
108
110
130
TRANSLATOR SERVER

FIG. 1

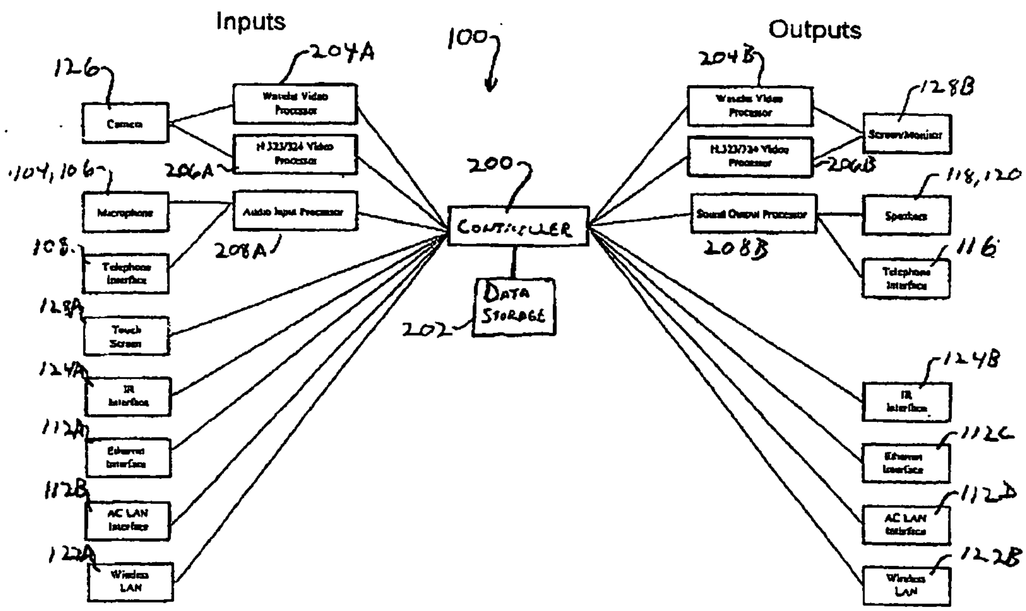

Inputs
Outputs
100
200
Controller
Data Storage
202
126
Camera
204A
Wavelet Video Processor
206A
H.323/324 Video Processor
104, 106
Microphone
208A
Audio Input Processor
108
Telephone Interface
128A
Touch Screen
124A
IR Interface
112A
Ethernet Interface
112B
AC LAN Interface
122A
Wireless LAN
204B
Wavelet Video Processor
206B
H.323/324 Video Processor
128B
Screen/Monitor
208B
Sound Output Processor
118, 120
Speakers
116
Telephone Interface
124B
IR Interface
112C
Ethernet Interface
112D
AC LAN Interface
122B
Wireless LAN

FIG. 2

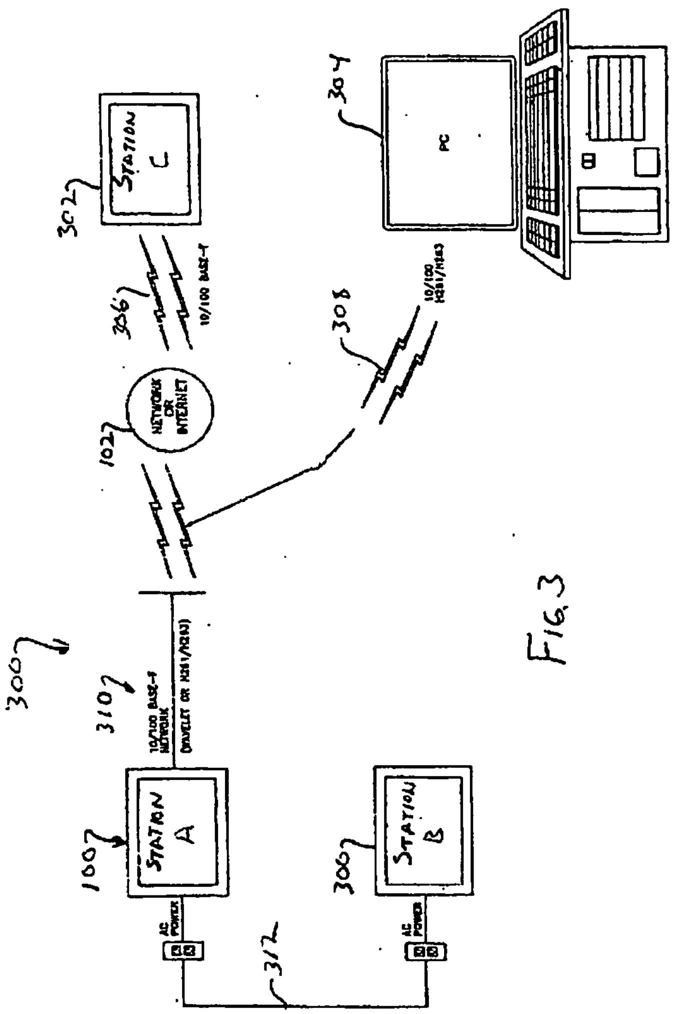

STATION C
302
306
10/100 BASE-T
NETWORK OR INTERNET
102
304
PC
308
10/100 H261/H263
300
310
10/100 BASE-T NETWORK
(WAVELET OR H261/H263)
STATION A
100
STATION B
306
AC POWER
AC POWER
312

FIG. 3

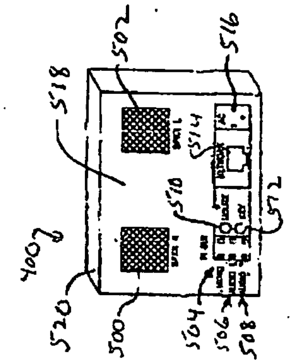

518
502
516
514
510
512
400
520
500
504
506
508

FIG. 5

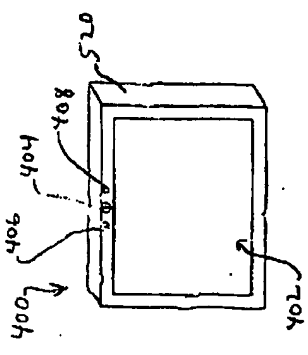

520
408
404
406
400
402

FIG. 4

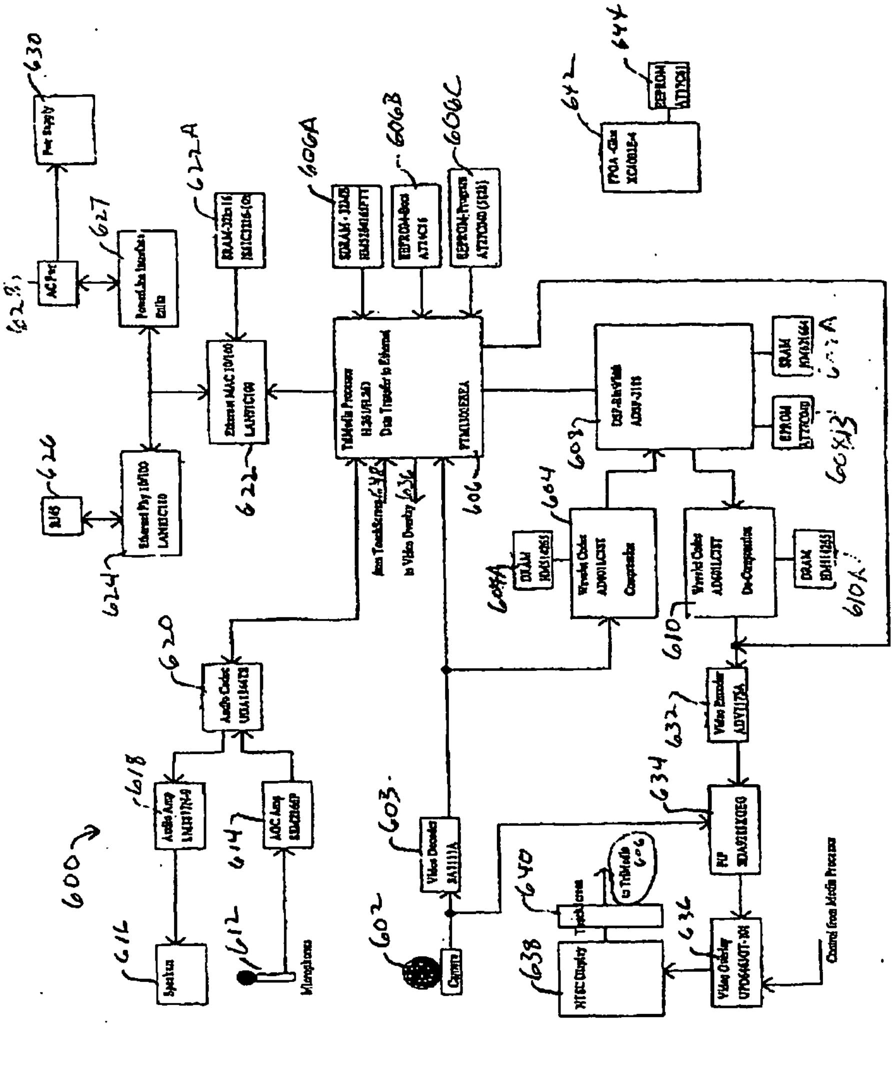


Fig. 6